

proyecto de proceso dE software

Creación de la nube de palabras



18 de diciembre de 2019

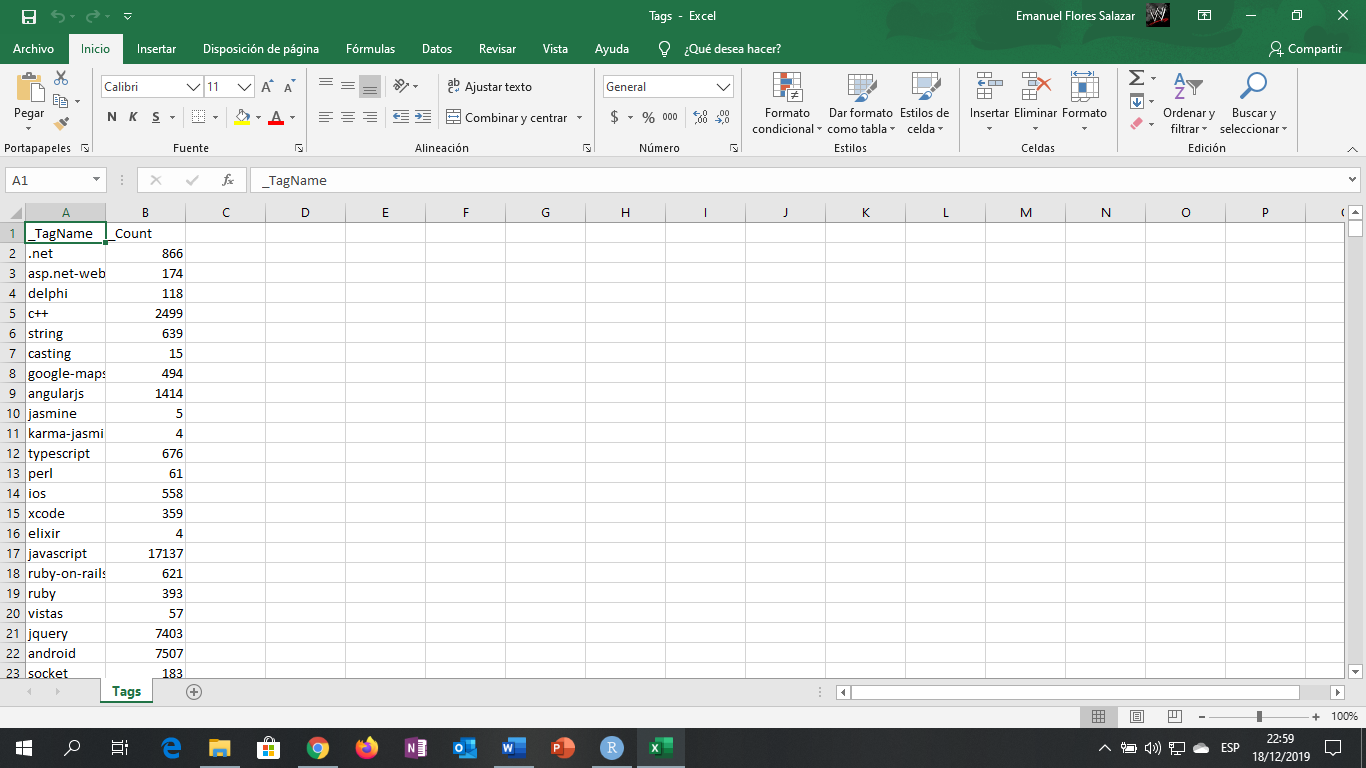
Universidad de guayaquil

Emanuel Ernesto Flores Salazar

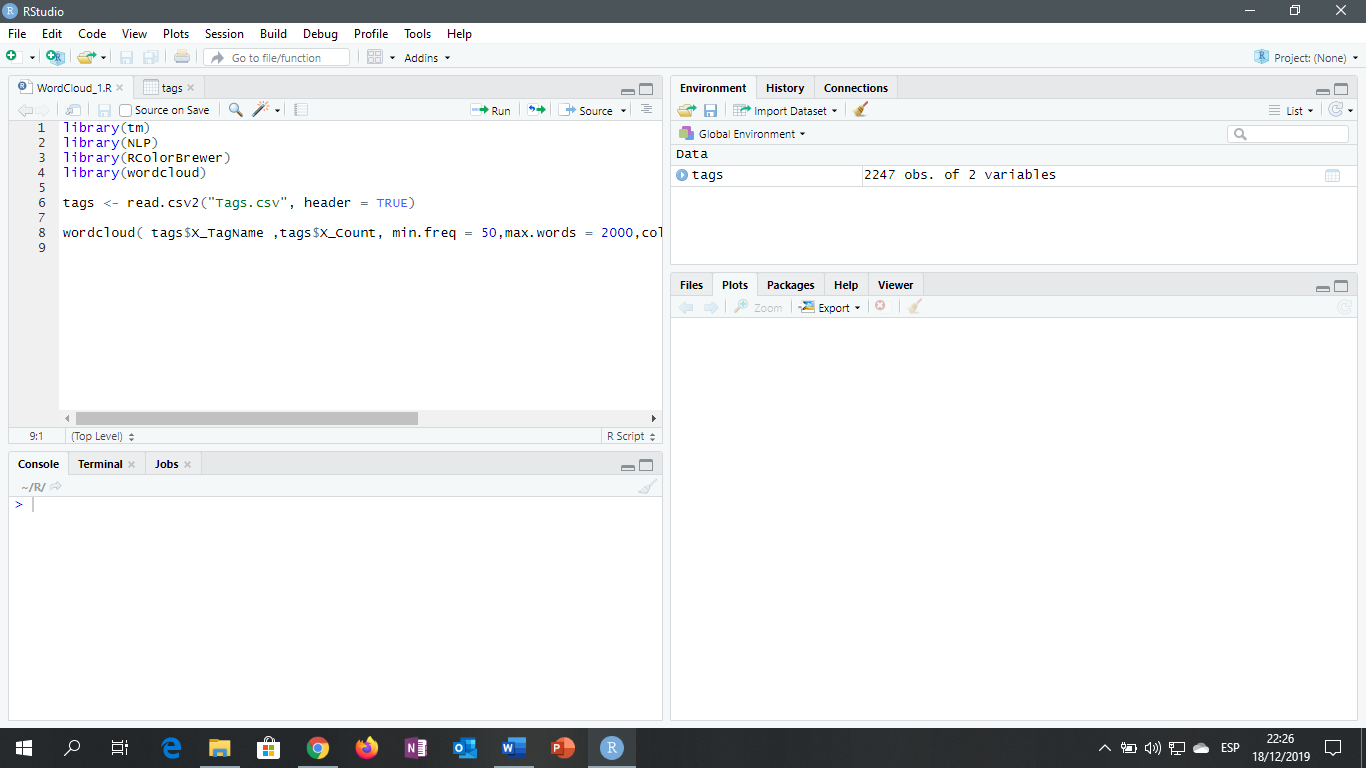
**Creación de la nube de palabras**

Para el proyecto, se utilizó lo siguiente:

1. RStudio (Software para Análisis Estadísticos).
2. Los datos que se obtuvieron en la plataforma Stack Exchange Data Dump, que contiene los Tags de los usuarios de Stack Overflow.
3. El documento originalmente era formato XML, y se lo convirtió a formato CSV y XLSX ya que se puede trabajar de manera organizada, entendible.



1. Este Excel nos sirvió de guía para poder crear la nube de palabras y se trabajó con las columnas TagName y Count ya que nos indican el nombre de las etiquetas y la frecuencia que se encuentra en la plataforma.
2. Una vez obtenida la tabla, ahora si podemos trabajar con el entorno de trabajo (RStudio).



1. A continuación, voy a definir cada línea de código en RStudio; lo primero serían las librerías:

*library(tm)* <- específico para minería de textos, análisis estadísticos.

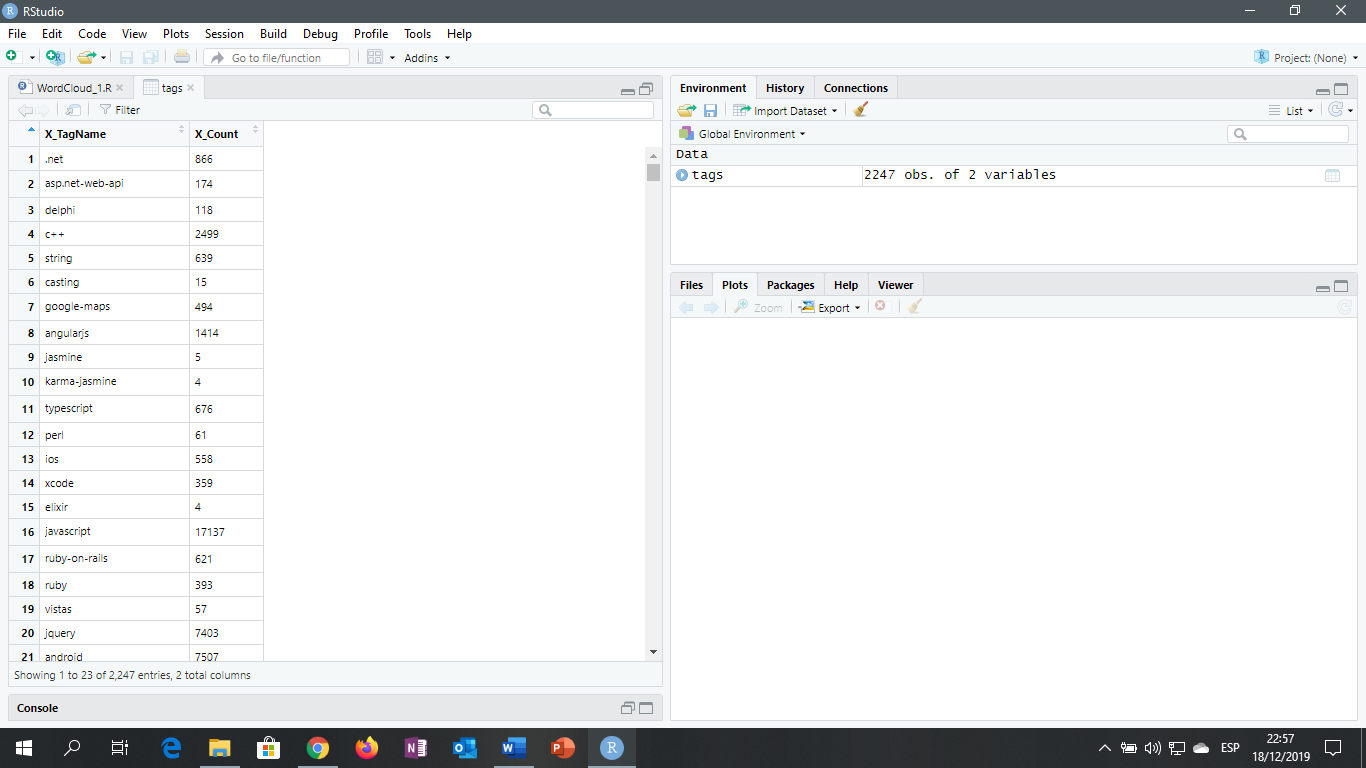
*library(NLP)* <- procesamiento del lenguaje natural (PNL), se usa para encontrar el componente del sistema bajo prueba con la mayoría de los problemas encontrados, pero no hace nada y si borramos esta librería, no afectaría en nada en nuestro código.

*library(RColorBrewer)* <- permite a los usuarios crear gráficos coloridos con paletas de colores prefabricadas que visualizan los datos de manera clara y distinguible.

*library(wordcloud)* <- para graficar la nube de palabras.

Lo siguiente será:

*tags <- read.csv2("Tags.csv", header = TRUE) <-* poner el nombre del archivo con el que vamos a trabajar, en este caso será tags donde contiene todos los nombres de las etiquetas y la frecuencia; lo siguiente es que después lo lea con la extensión “.csv2” para que se haga solamente 2 columnas (nombres de las etiquetas y la frecuencia) y después de eso dentro de los paréntesis se debe poner el archivo que anteriormente era xml pero lo convertimos en csv y lo guardamos como Tags con el tipo “Archivo de valores separados por comas de Microsoft Excel” para que al momento de ejecutar esta línea de código, se pueda observar la tabla de tags.



Y, por último:

*wordcloud( tags$X\_TagName ,tags$X\_Count, min.freq = 50,max.words = 2000,colors = brewer.pal(8,"Dark2"),random.order = FALSE)* <- vamos a crear a crear la nube de palabras con el mismo nombre de la librería, dentro de los paréntesis ponemos el nombre del archivo que en este caso es tags y a continuación el nombre de la columna que vendría ser X\_TagName, así mismo con el otro X\_Count; lo siguiente es min.freq que en este caso solo puede los valores tomar a partir de 50 o superior, max.words que es el número máximo de palabras que se trazarán. términos menos frecuentes caídos, en este caso es 2000, colors=brewel.pal hace que las paletas de colores de ColorBrewer estén disponibles como paletas y le puse (8,"Dark2") y el random.orden que no importa el orden aleatorio al momento en el que se ejecute el programa por eso es FALSE.

1. El resultado final de la creación de la nube de palabras que vemos a continuación:

